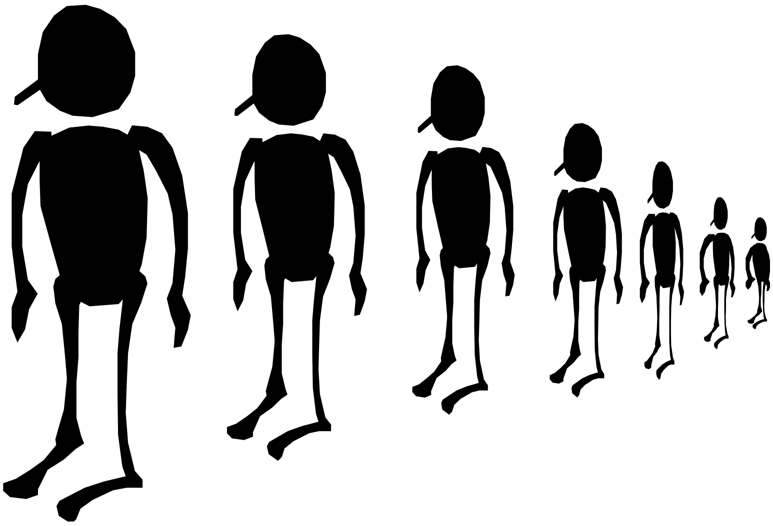
Universidade de Cabo Verde

Teoria das filas de espera



Autor: Ivanildo Cabral Moreira

Setembro de 2009

Uni-Cv

DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA

LICENCIATURA EM ENSINO DE MATEMÁTICA

Trabalho Científico De Fim de Curso

Tema:

Teoria das Filas de Espera

O júri

____

____

____

Setembro 2009

Ivanildo Cabral Moreira

Dedicatória

O autor dedica toda a sua inquietação, o seu esforço e o seu desempenho exclusivamente à sua família. Espera-se que o entusiasmo, seriedade e empenho desse trabalho sirva-lhe de estímulo para fazer sempre "mais e melhor".

Agradecimentos

Embora um trabalho científico tenha finalidade académica, há contributos de naturezas diversas que não podem nem devem deixar de ser realçados. Por essa razão, o autor expressa, com todo carinho e respeito, os seus sinceros agradecimentos:

- ✓ Primeiramente pelo criador do universo que sempre lhe proporcionou energia e fé, principalmente nessa pesquisa;

- ✓ Aos seus pais: Casimiro Silva Moreira e Hilária Mendes Cabral pelo grande esforço e coragem que têm feito para educar os filhos e acompanhá-los nos estudos; pelo estímulo e apoio incondicional desde a primeira hora de vida; pela paciência e grande amizade com que sempre o ouviram; pela sensatez com que sempre o ajudaram; pela dedicação que tiveram para acolher o seu filho que, embora longe dos pais, teve uma boa educação;

- ✓ Às suas irmãs Janice, Vanice e Milena pela amizade, pela orientação, confiança; pelo apoio; pela motivação e coragem. Não tem palavras que manifesta toda a sua emoção de pertencer essa brilhante família;

- ✓ Ao seu avô José e sua avó Antónia por tudo que o ensinaram desde o berço;

- ✓ Aos seus tios Pedro e António que o acolheu em sua casa na praia durante o tempo de estudo secundário e superior, também não menos importante aos seus primos António Pedro, Carla Elisabete e Manuela Moreira;

- ✓ À sua namorada Neusa Fernandes pelo amor carinho e dedicação; pela grande paciência que ela teve de aceitar as descargas da sua

saturação, pois ela viveu com ele todos os momentos difíceis sem nenhuma reclamação e sempre tentou acatar os seus desalentos com toda ternura;

- Aos seus grandes amigos Valdimiro da Moura e Adilson Semedo que, como irmãos, partilharam consigo um espaço onde sempre pôde procurar o conforto;

- A todos os seus familiares amigos e colegas de formação em especial Victor Vieira e Silvino dos Reis pelos bons momentos e momentos ruins que passaram e às restantes pessoas que directa ou indirectamente contribuíram para que este trabalha se concretizasse.

- Ao seu orientador pela humildade e pelo grande contributo proporcionado a esta pesquisa;

A todos, os seus sinceros agradecimentos e muito obrigado!

O autor: Ivanildo Cabral Moreira

O orientador: Professor Dr. Adilson Barros

Índice

CAPÍTULO 1

Introdução

As …las de espera estão presentes em muitas das situações do quotidiano, como por exemplo: …las de espera de um banco, …las em uma clínica, …las em o…cinas, …las em processos industriais de produção, …las em redes de computadores, etc.

A teoria das …las é um ramo da probabilidade que estuda a formação de …las, através de análises matemáticas precisas e propriedades mensuráveis das …las. Ela provê modelos para demonstrar previamente o comportamento de um sistema que ofereça serviços cuja demanda cresce aleatoriamente, tornando possível dimensioná-lo de forma a satisfazer os clientes e ser viável economicamente para o provedor do serviço, evitando desperdícios e gargalos.

O uso da Teoria de Filas, na prática, envolve dois maiores aspectos: selecção de um modelo matemático apropriado que represente adequadamente o sistema, com o objectivo de determinar as medidas de desempenho mais adequadas e a implementação de um modelo de decisão baseado nas medidas de desempenho, com o objectivo de melhorar o processamento do serviço ao cliente, a despeito das di…culdades matemáticas.

O cenário de teoria das …las pode ser descrito do seguinte modo: um recurso (também chamado de centro de serviço) é compartilhado por um conjunto de clientes que de tempos em tempos vão a esse recurso para receberem seu serviço. Se ao solicitar o serviço o cliente encontrar o recurso disponível, ele será imediatamente atendido. Caso contrário, se o recurso já estiver atendendo outro(s) cliente(s) no momento da solicitação, ele deverá aguardar em uma …la.

Pretendemos desenvolver uma investigação científica, de base sobretudo bibliográfica, sobre a Teoria das Filas e alcançar os seguintes objectivos:

- Iniciar ao estudo da teoria das filas;
- Construir um instrumento de referência para o conhecimento da teoria das filas;
- Treinar em matemática aplicada, como instrumento de modelação e de ensino (em especial no campo da probabilidade).

***

O presente trabalho encontra-se dividido em dois capítulos, no primeiro falamos da abordagem histórica, da descrição de um sistema de filas, das distribuições de probabilidades, do processo estocástico, do processo de markov e também falamos do processo de nascimento e morte.

Já no segundo capítulo, optamos por falar do Desempenho dos Modelos, onde abordamos a famosa lei de little; os modelos makovianos, não markovianos e em série e também o comportamento transitório de um sistema de filas. Finalmente apresentamos algumas considerações finais e um anexo.

CAPÍTULO 2

Teoria das Filas

2.1 Fundamentação teórica

2.1.1 Abordagem Histórica das Filas de Espera

Agner Krarkup Erlang (1 de Janeiro de 1878 a 3 de Fevereiro de 1929) foi o matemático, estatístico, engenheiro dinamarquês que idealizou pela primeira vez os conceitos de Engenharia de Tráfego e de Teoria das Filas.

Trabalhando na empresa “Copenhagen Telephone Company”, foi que Erlang teve que resolver um clássico problema de determinar quantos circuitos são necessários para providenciar um atendimento aceitável nas chamadas telefónicas. Mas seu raciocínio o ajudou a perceber que a matemática resolveria outro problema, isto é: quantos operadores de telefone são necessários para atender um número de chamadas telefónicas determinadas previamente.

Nessa época a maioria das centrais telefónicas, usava trabalhadores como operadores para gerenciar as chamadas telefónicas, conectando os …os telefónicos nas tomadas eléctricas das placas com circuitos.

Existiam avanços nas aplicações telefónicas, mas na teoria das …las não teve um avanço semelhante. Logo, a partir da década dos anos 50, foi quando as aplicações em áreas além dos sistemas de telefone começaram a evoluir.

Erlang trabalhou no desenvolvimento da área de tráfego nos sistemas de chamadas telefónicas e publicou o seguinte: Em 1909, “A Teoria das Probabilidades e as Conversações Telefónicas” em que provou que a distribuição de Poisson se aplica ao tráfego aleatório de chamadas telefónicas. Em 1917, “Soluções de Alguns Problemas na Teoria de Probabilidades de Importância nas Chamadas Automáticas de Telefone" em que inclui sua fórmula clássica de tempo de espera e tempo perdidos.

2.1.2 Descrição das Distribuições de Probabilidade

Distribuição de Poisson

A distribuição de Poisson, cujo nome se deve ao físico Francês Simon Poisson (1781-1840), permite descrever uma grande variedade de situações com aplicações em muitas áreas do conhecimento.

Definição 2.1 *Seja N uma variável aleatória discreta, tomando os seguintes valores:* $0, 1, 2, \ldots$ *quando*

$$P(N = n) = \frac{e^{-\alpha}\alpha^n}{n!}, (n = 0, 1, 2, \ldots)$$

N tem distribuição de Poisson, com parâmetro $\alpha > 0$.

Para verificar que a expressão acima representa uma distribuição de probabilidade, basta observar que:

$$\sum_{n=0} P(N = n) = \sum_{n=0} \left(\frac{e^{-\alpha}\alpha^n}{n!} \right) = e^{-\alpha} \times e^{\alpha} = 1$$

onde

$$\sum_{n=0} \frac{\alpha^n}{n!} = 1 + \alpha + \frac{\alpha^2}{2!} + \ldots + \frac{\alpha^n}{n!} = e^{\alpha}$$

Valor Esperado Se N tiver distribuição de Poisson com parâmetro α então o valor esperado de N é

$$E(N) = \alpha.$$

Este valor esperado indica a média do número de eventos num intervalo de tempo.

Distribuição Exponencial

A distribuição exponencial está intimamente relacionada com o processo de Poisson. Demonstra-se que, num processo de Poisson, o "tempo" de espera até ao primeiro sucesso (ou entre dois sucessos consecutivos) segue uma certa distribuição exponencial.

Definição 2.2 *Uma variável aleatória contínua T, que assume todos os valores não negativos, terá uma distribuição exponencial com parâmetro $\alpha \geq 0$, se sua função densidade de probabilidade (fdp) for dada por:*

$$f(t) = \begin{cases} \alpha e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

Demonstração

Por definição a probabilidade de uma variável aleatória contínua é sempre igual à área do gráfico da curva da fdp. Neste caso:

$$p(t \geq 0) = \int_0^{\infty} f(t)\, dt = \int_0^{\infty} \alpha e^{-\alpha t} dt = -\int_0^{\infty} (-\alpha) e^{-\alpha t} dt$$

$$\int e^{f(t)} f'(t) dt = e^{f(t)} + c, \text{ onde } c \text{ é uma constante qualquer}$$

Então:

$$-\int_0^{\infty} (-\alpha) e^{-\alpha t} dt = -e^{-\alpha t}\Big|_0^{\infty} = -e^{-\infty} + e^0 = 0 + 1 = 1$$

como a probabilidade acumulada

$$P(T \geq t) = \int_t^{\infty} \alpha e^{-\alpha t} dt = -e^{-\alpha t}\Big|_t^{\infty} = -e^{-\alpha \infty} + e^{-\alpha t} = 0 + e^{-\alpha t} = e^{-\alpha t}$$

também

$$P(T \geq t) = e^{-\alpha t}$$

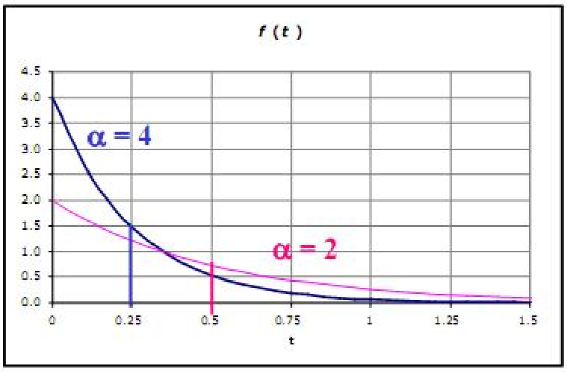
Esperança Matematica Indica a média do tempo de duração dos eventos descritos pela distribuição exponencial. Seu valor é

$$E(T) = \frac{1}{\alpha}$$

Variânçia A variância é igual ao quadrado do valor esperado, ou seja,

$$var(T) = \frac{1}{\alpha^2}.$$

A distribuição exponencial é adequada para tempos de serviço quando este é em geral muito curto e ocasionalmente muito longo. Como por exemplo, nos Bancos de hospitais, Bancos, lojas, etc.



Propriedade 1: A função densidade de probabilidade $[f(t)]$é uma função estritamente decrescente em t A distribuição exponencial é adequada para tempos entre chegadas, em situações em que potenciais clientes desistem (e voltam mais tarde) quando outro cliente já está na fila. Desta forma, vão aparecendo mais ou menos regularmente (curtos intervalos) com intervalos ocasionalmente longos sem aparecer nenhum cliente.

Propriedade 2: Propriedade da não-memória (falta de memória) da distribuição exponencial

$$P[T > t + \Delta t \mid T > \Delta t] = P[T > t] \text{ para } t > 0 \text{ e } \Delta t > 0$$

Demonstração:

$$
\begin{aligned}
P[T > t + \Delta t \mid T > \Delta t] &= \frac{P[T > \Delta t, T > t + \Delta t]}{P[T > \Delta t]} = \\
&= \frac{P[T > t + \Delta t]}{P[T > \Delta t]} = \frac{e^{-\alpha(t+t\Delta)}}{e^{-\alpha \Delta t}} = e^{-\alpha t} = P[T > t]
\end{aligned}
$$

A distribuição de probabilidades do restante tempo até ao próximo evento (chegada de um novo cliente) é a mesma independentemente de há quanto tempo ocorreu o último evento (chegada do último cliente). Assim, o tempo entre chegadas é independente de quando aconteceu a última chegada e tem-se situações com diferentes tempos de serviço.

Propriedade 3: O mínimo de variáveis exponenciais independentes é uma distribuição exponencial Sejam $T_1, T_2, ..., T_n$ variáveis aleatórias independentes com distribuições exponenciais de parâmetros $\alpha_1, \alpha_2, ..., \alpha_n$.

Seja U uma v.a., onde:

$$U = \min \{T_1, T_2, ..., T_n\}$$

Se T_i representa o instante em que ocorre um destes eventos, então U representa o instante em que o primeiro dos n eventos ocorre.

$$
\begin{aligned}
P[U > t] &= P[T_1 > t, T_2 > t, ..., T_n > t] = \\
&= P[T_1 > t] P[T_2 > t] ... P[T_n > t] = \\
&= e^{-\alpha_1 t} e^{-\alpha_2 t} ... e^{-\alpha_n t} = \\
&= e^{-\left(\sum_{i=1} \alpha_i\right) t}
\end{aligned}
$$

Para o tempo entre chegadas, consideramos que existem n tipos de clientes diferentes com diferentes distribuições exponenciais com $\alpha_1, \alpha_2, ..., \alpha_n$.Segundo P 2 (Propriedade da não memória), o tempo que falta, a partir de um dado instante, até à chegada de um cliente de tipo i tem também uma exponencial de parâmetro α_i (mesma distribuição) e segundo P3 o tempo que falta, a partir de um dado instante, até à chegada de um cliente de qualquer tipo tem também uma exponencial de parâmetro:

$$\alpha_U = \sum_{i=1} \alpha_i$$

Para o tempo de serviço, assumimos que existem s servidores em paralelo com a mesma distribuição exponencial (com parâmetro μ) dos tempo de serviço.

Se T_i é o tempo de serviço que ainda falta, a partir de um dado instante, para o servidor i, então a distribuição de probabilidades do tempo até que um próximo servidor termine o serviço é uma exponencial com parâmetro $n\mu$.

Ou seja o sistema multi-servidor pode ser visto como um sistema mono-servidor cuja distribuição do tempo de serviço é $n\mu$.

Propriedade 4: Relação com a distribuição de Poisson

- Seja $N(t)$ o número de ocorrências de um evento no intervalo de tempo entre 0 e t ($t \geq 0$) uma variável aleatória com a seguinte distribuição de probabilidade

$$P[N(t) = n] = \frac{(\alpha t)^n e^{-\alpha t}}{n!} \text{ para } n = 0, 1, 2, \ldots$$

- $N(t)$ tem uma distribuição de Poisson com parâmetro αt.
- A correspondente esperança matemática é:

$$E(N(t)) = \alpha t$$

Então, o número esperado de eventos por unidade de tempo é α (α é designado de ritmo médio de ocorrência de eventos).

- Com $n = 0$ tem-se:

$$P[N(t) = 0] = e^{-\alpha t}$$

que é a probabilidade de que o primeiro evento ocorra depois do tempo t. Trata-se de uma distribuição exponencial de probabilidade sobre t.

- Quando os eventos são controlados numa base contínua, o processo contínuo $\{N(t); t \geq 0\}$ é designado de Processo de Poisson.

- Se os tempos de serviço seguem uma distribuição exponencial de parâmetro μ então define-se $N(t)$ como o número de serviços concluídos por um servidor continuamente ocupado durante um tempo t, com $\alpha = \mu$.

- Para modelos multi-servidores o número de serviços concluídos por n servidores continuamente ocupados durante um tempo t, com

$$\alpha = n\mu.$$

- Se os tempos entre chegadas de novos clientes seguem uma distribuição exponencial de parâmetro λ então definimos $N(t)$ como sendo o número de chegadas durante um tempo t, com $\alpha = \lambda$ (que é o ritmo médio de chegadas).

Distribuição de Erlang

Se os tempos de atendimento não obdecem a uma distribuição exponencial, então eles são modelados por uma distribuição de Erlang. A distribuição de Erlang é uma variável randômica continua com parâmetro de razão R e parâmetro k. O parâmetro k tem que ser inteiro ≥ 1 e também controla a forma do gráfico da função densidade de probabilidade (fdp) Erlang. A função densidade de probabilidade é dada do seguinte modo:

$$f(t) = \frac{R(Rt)^{k-1}e^{-Rt}}{(k-1)!} \qquad (t \geq 0, k \geq 1), R = k\mu$$

Fazendo $k = 1$, podemos notar que $f(t)$ se converte na distribuição exponencial com R.

$$f(t) = R \times e^{-Rt}$$

Como $R = k\mu$ e $k = 1$, tem-se $R = \mu$, daí:

$$f(t) = \mu e^{-\mu t}$$

A distribuição Erlang é igual ao somatório de um número k de variáveis randômicas exponenciais independentes, ou seja, $E_k = T_1 + T_2 + ... + T_k$, onde T_i é uma variável randômica exponencial com parâmetro $R = k\mu$. O número k é chamado número das fases da distribuição Erlang.

Valor Esperado (média)

$$E(T) = \frac{1}{\mu}$$

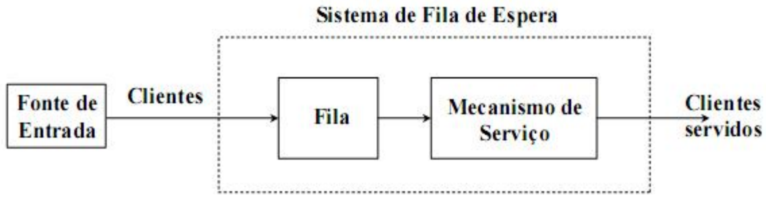
Variância

$$var(t) = \frac{k}{R^2} = \frac{k}{(k\mu)^2} = \frac{1}{k\mu^2}.$$

se k ¡ ! 1 logo $var(t)$ ¡ ! 0.

2.1.3 Descrição de um Sistema de Filas

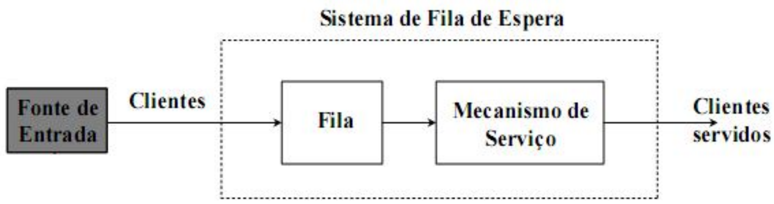
Processo básico de Fila de Espera



Uma fonte de entrada gera ao longo do tempo clientes que solicitam um serviço.

Os clientes por sua vez entram no Sistema de Fila de Espera e juntam-se a uma Fila de Espera e em certos instantes é escolhido um membro da …la, para ser servido, de acordo com alguma regra conhecida por Disciplina da Fila de Espera de seguida o cliente seleccionado é servido por um mecanismo de serviço, posteriormente, quando um serviço é concluído para um cliente, este sai do sistema de Fila de Espera.

Fonte de entrada -População alvo



[2] A dimensão da fonte constitui o número total de clientes que podem requerer serviços do sistema, este pode ser …nito ou in…nito.

– In…nito: a fonte é ilimitada.

- Os cálculos são mais simples;
- Também é assumido quando a dimensão é …nita mas grande;

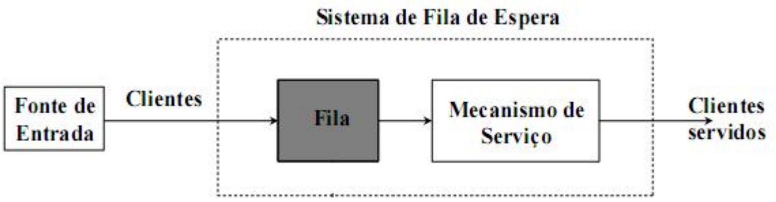
– Finito: a fonte é limitada.

- Os modelo analítico já são mais complicado pois o número de clientes dentro do sistema (na …la ou a ser servidos)afecta o número de clientes fora do sistema.
- Este modelo deve ser adoptado sempre que o ritmo a que os clientes são gerados pela fonte depende signi…cativamente do número de clientes que estão dentro do sistema.

² O Padrão estatístico segundo o qual os clientes se apresentam para serem servidos, pode seguir uma distribuição de Poisson. Assume-se que a chegada de clientes ao sistema é independente do número de clientes presentes (população in…nita).

² O tempo entre chegadas de clientes ao sistema segue uma distribuição exponencial.

Fila de Espera



² A dimensão da …la de Espera pode ser tanto …nita como in…nita.

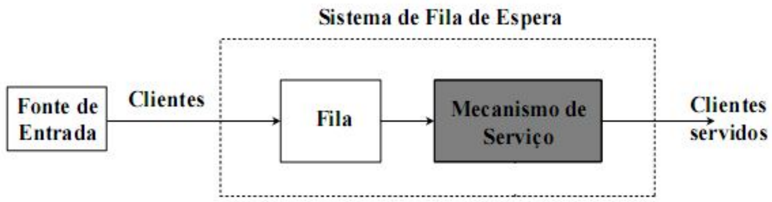
– A suposição de …la de capacidade in…nita é a forma mais geral, mesmo quando a capacidade for …nita mas su…cientemente grande.

– Quando o limite é …nito e pequeno de tal modo que a capacidade da …la possa ser atingida com frequência, então assume-se que a capacidade é um número …nito. Assim, os modelo são mais complexos.

² A Disciplina da Fila de Espera descreve a forma como os clientes saem da …la de espera para serem atendidos. Podem ter as seguintes disciplinas:

– Primeiro a chegar, primeiro a ser servido;
– Aleatório;
– Prioridades, etc.

Mecanismo de Serviço



² A organização pode ter uma ou mais infra-estruturas de serviço. Se for mais de uma, cada cliente deve ser servido sequencialmente por todas elas. E cada infra-estrutura de serviço, por sua vez, é composta por um ou mais servidores em paralelo.

² Para cada servidor é necessário especi…car a distribuição de probabilidades dos tempos de serviço (eventualmente um por cada tipo de cliente). Geralmente, todos os servidores têm a mesma distribuição de Probabilidades.

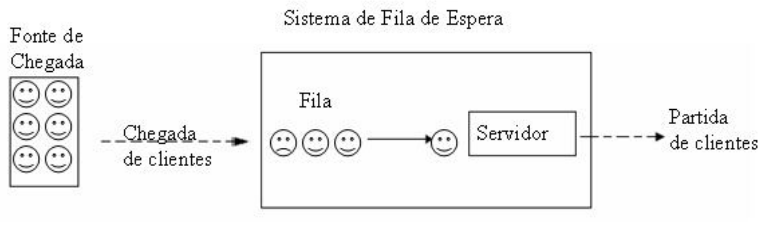
² As distribuições de probabilidades comuns são:

– Distribuição Exponencial;

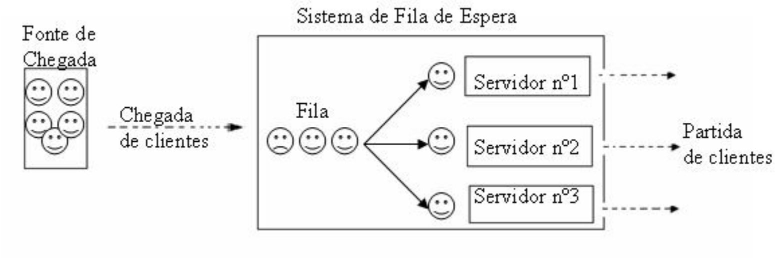
– Distribuição degenerada (constante);

– Distribuição de Erlang.

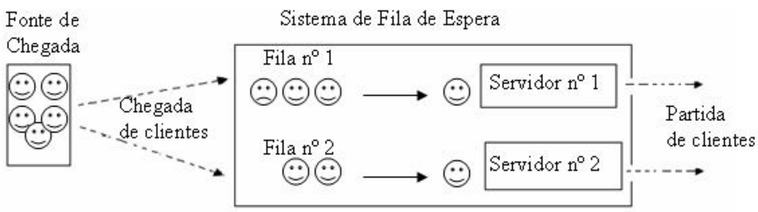
Con…guração da Fila de Espera:



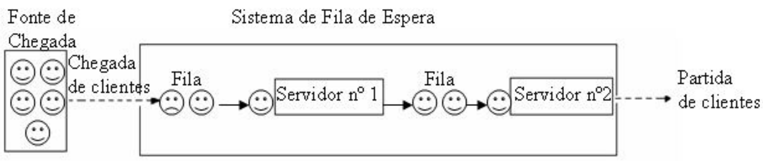
(a) Única fila de espera, único servidor



(b) Única fila de espera, múltiplos servidores em paralelo



(c) Múltiplas filas de espera, múltiplos servidores em paralelo



(d) Única fila de espera, múltiplos servidores em série

Exercício 2.1 *Identi...que os clientes, os servidores e as características da ...la que se forma, num posto de lavagem automática de automóveis, com uma única via de acesso.*

Os clientes são os automóveis que entram no local para serem lavados. O servidor é a máquina que faz a lavagem e a via única indica um ou mais servidores em série. Geralmente, a lavagem de automóveis opera de modo a que o primeiro a chegar é também o primeiro a ser atendido. Assim a disciplina da ...la de espera é FIFO. A capacidade do sistema é o número de carros que podem manobrar em segurança no local da lavagem automática. Se for permitido que os automóveis possam ...car estacionados na via plublica para entrada posterior no local da lavagem automática, então o sistema tem capacidade in...nita.

Terminologia e Notação

O Estado do Sistema é constituido pelo número de clientes dentro do sistema de …la de espera (na …la ou a ser servido pelos servidores).

O comprimento da …la é número de clientes na …la à espera de serviço. Este é igual ao estado do sistema menos o número de clientes a serem servidos.

Denotamos o número de clientes no sistema no instante t (t ¸ 0) por $N(t)$, a probabilidade de estarem exactamente n clientes no

sistema no instante t, conhecido o número de clientes no instante $t = 0$ por $P_n(t)$, o número de servidores (canais paralelos) no sistema por s, o ritmo médio de chegadas de novos clientes quando estão n clientes no sistema (número esperado de chegadas por unidade de tempo) por λ_n.

Se λ_n é constante para todos os valores de n, ou seja, quando o ritmo de chegada não depende do número de clientes no sistema, denota-se por λ.

E $\frac{1}{\lambda}$ o tempo esperado entre chegadas de novos clientes.

O ritmo médio de serviço global do sistema (número médio de clientes que terminam seu serviço por unidade de tempo) denota-se por μ_n.

Nota 2.1 *μ_n é um valor combinado do ritmo de serviço de todos os servidores ocupados.*

Denotamos por μ quando o ritmo médio de serviço, por servidor ocupado, é constante para todos, assim, $\mu_n = s\mu$ quando n ¸ s, isto é, quando todos os servidores estão ocupados e $\frac{1}{\lambda}$ o tempo esperado de serviço.

O factor (taxa) de utilização da infra-estrutura de serviço é

$$\rho = \frac{\lambda}{s\mu}$$

Isto é, a fracção de tempo esperado em que os servidores estão ocupados, onde μ é o número médio de clientes que terminam o seu serviço por unidade de tempo por servidor sempre ocupado, s é o número de servidores, $s\mu$ é o número médio de clientes que terminam o seu serviço por unidade de tempo supondo que todos os servidores estão ocupados, ou seja, é a capacidade de serviço do sistema por unidade de tempo e λ é o número esperado de novos clientes por unidade de tempo.

Terminologia e Notação: Regime Estacionário

Para sistema em regime estacionário, a distribuição de probabilidade do sistema mantém-se constante ao longo do tempo.

As grandezas de...nidas para o sistema em regime estacionário são:

– P_n - probabilidade de estarem exactamente n clientes no sistema;

– L - número esperado de clientes no sistema;

– L_q - comprimento esperado da ...la de espera (excluindo os clientes que estão a ser servidos);

– W - tempo passado no sistema (incluindo o tempo de serviço) para cada cliente;

$$W = E(W)$$

– W_q - tempo de espera no sistema (excluindo o tempo de serviço) de cada cliente;

$$W_q = E(W_q)$$

Também são estabelicidas algumas relações entre L, W, L_q e W_q, onde assumindo que λ_n, ritmo médio de chegadas ao sistema é constante e igual a λ para todo n, veri...ca-se num regime estacionário que:

$$L = \lambda \text{ £ } W \text{ (Fórmula de Little)}$$

e

$$L_q = \lambda \text{ £ } W_q$$

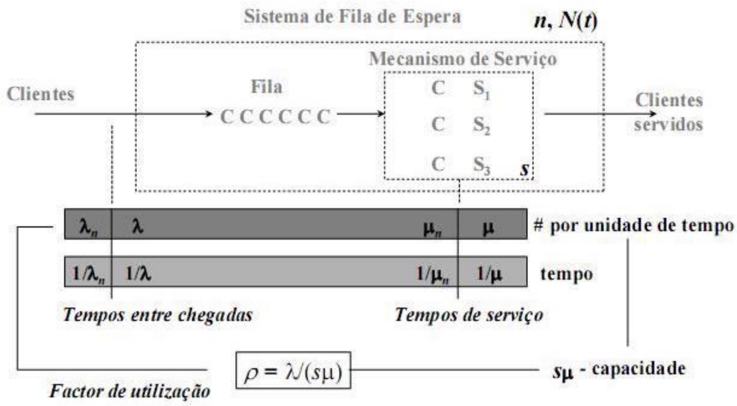
Nota 2.2 *Se λ_n não toma o mesmo valor para todos os valores de n, então, é possível substituir λ por $\overline{\lambda}$ que é o valor médio dos λ_n ao longo do tempo.*

Por outro lado, assumindo que o tempo médio de serviço é um valor constante e igual a $\frac{1}{\mu}$ para n ¸ 1, então:

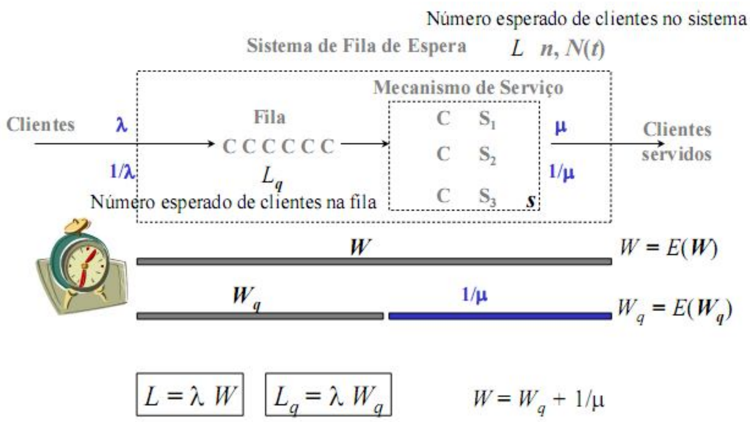
$$W = W_q + \frac{1}{\mu}$$

Ou seja, o tempo passado no sistema é igual ao tempo passado à espera mais o tempo de serviço.

Resumo da terminologia Tomando como referência um sistema de …la de espera com uma única …la e multiplos servidores em paralelo, podemos resumir da seguinte forma:



E para um sistema em regime estacionário, pode ser:



2.1.4 Processo Estocástico

De…nição 2.3 *Um processo estocástico é uma variável que se desenvolve no tempo de uma maneira que é pelo menos parcialmente aleatória e imprevisível. De uma maneira mais formal, um processo estocástico é de…nido por* *uma lei de probabilidade para a evolução de uma variável x durante um tempo t.*

Suponhamos que existam apenas 3 períodos, ou seja, $T = \{1, 2, 3\}$ e que no período 1 a variável aliatória x_1 tenha uma distribuição Normal, enquanto no período 2 a variável x_2 tenha uma distribuição t de Student e no período 3 a variável x_3 tenha uma distribuição Chi-Quadrado. Logo, essa sequência de três variáveis aleatórias é na verdade um processo estocástico. Também um processo estocástico pode ser formado por variáveis identicamente distribuídas, ou seja, nesse exemplo poderíamos ter que todas as três variáveis tivessem distribuição Normal.

Os processos estocásticos podem ser classificados da seguinte maneira:

- Processos Estacionários: as propriedades estatísticas (média e variância) desta variável são constantes no tempo. Ex: temperatura do rio. Num processo estritamente estacionário, além da média e da variância, todos os seus demais momentos também devem ser necessariamente constantes.
- Processos não-estacionários: o valor esperado da variável aleatória pode crescer sem limite e sua variância, T anos a frente, aumenta com T. Ex: Acções.
- Processos de Estado Discreto: A variável objecto pode assumir somente alguns valores discretos. Ex: Taxas de juros Selic. (variações mínimas de $0,25\%$).
- Processos de Estado Contínuo: A variável objecto pode assumir qualquer valor. Ex: Temperatura.
- Processos de Tempo Discreto: t é uma variável que assume apenas valores discretos. O valor da variável objecto somente pode mudar apenas nestes pontos. Ex: Taxa Selic – muda apenas em determinados intervalos, nunca todo dia.
- Processos de Tempo Contínuo: t é uma variável contínua. Ex: temperatura.

Em muitas situações práticas, os atributos de um sistema mudam de forma aleatória com o tempo, como por exemplo: o número de clientes numa fila de espera, o congestionamento no trânsito, o número de itens num depósito, ou o valor de uma acção financeira, entre outras. Em algumas circunstâncias, é possível descrever os fundamentos do processo que explica como a mudança ocorre. Quando as características do processo são governadas pela teoria da probabilidade, se tem um processo estocástico.

O primeiro passo para modelar um processo dinâmico é de…nir o conjunto de estados que pode alcançar e descrever os mecanismos que governam suas transições.

Um estado é como um *snapshot* (foto intantânea) do sistema em um tempo determinado. É uma abstração da realidade que descreve os atributos de um sistema que interessa.

O tempo é uma medida linear através da qual o sistema se movimenta, e pode ser visto como um parâmetro. Devido à existência do tempo, existe: passado, presente e futuro.

Usualmente se sabe qual foi a trajectória que o sistema tomou para chegar ao estado actual. Usando esta informação, o objectivo é antecipar o futuro comportamento do sistema em termos básicos de um conjunto de atributos.

Por razões de modelagem, o estado e o tempo podem ser tratados de forma contínua e discreta. Mas, por razões computacionais e considerações teóricas, o estado somente será considerado em forma discreta. O tempo terá forma contínua ou discreta.

Para obter uma computação tratável, assumir que o processo estocástico satisfaz a propriedade de Markov, isto é, o caminho que o processo segue no futuro depende só do estado actual e não da sequência de estados visitados previamente ao estado actual.

Um tempo discreto no sistema induze ao modelo das Cadeias de Markov.

Para um tempo contínuo no sistema existe um modelo denominado de Processo de Markov.

Um modelo de um processo estocástico descreve actividades que terminam em eventos. Os eventos geram a transição de um estado a outro. Assumindo que a duração de uma actividade é uma variável aleatória contínua, eventos ocorrem na continuidade do tempo.

Nomenclatura de um Processo Estocástico

Dada uma variável aleatória, f$N(t)$g, onde t é um índice de tempo que toma valores de um conjunto dado T. T pode ser discreto ou continuo. $N(t)$ é escalar que pode assumir valores discretos ou contínuos. Considera-se somente processos estocásticos discretos …nitos.

O Tempo é o parâmetro de um processo estocástico.

O Estado é um vector que descreve atributos de um sistema em um tempo qualquer.

O vector estado tem m componentes.

$$S = (s_1, s_2, ..., s_m,).$$

$N(t)$ descreve algum atributo do estado.

O Conjunto de Estados é uma colecção de todos os estados possíveis.

Uma Actividade começa em um tempo determinado, tem uma duraçao e termina em um evento. Geralmente a duração de uma actividade é uma variável aleatória com uma distribuição de probabilidade conhecida.

Evento é a finalização de uma actividade. Um evento tem o potencial de mudar o estado do processo.

Calendário é o conjunto de eventos que podem ocorrer em um estado dado, $Y(s)$.

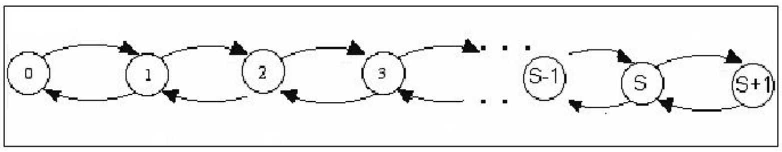
Próximo evento — Num estado qualquer, um ou mais eventos podem ocorrer. O próximo que ocorre é chamado de próximo evento. Começando pelo tempo actual, o tempo do próximo evento é o tempo mínimo que em termos matemáticos ficaria assim:

$$\delta_x = Minimun \{\delta_e | e \in Y(s)\}$$

O próximo evento é o valor de x que corresponde ao tempo mínimo. Quando as durações de eventos são variáveis aleatórias, ambos, o próximo evento e o tempo do próximo evento, são variáveis aleatórias.

Transição é a função que determina o próximo estado, s' baseado no estado actual s, e o evento, x. O número de elementos da função de transição é o mesmo do número de elementos do vector estado. $s' = T(s, x)$.

Diagrama de Transição de Estados é uma rede orientada na qual os estados são representados por *nós,* e eventos, representados por *arcos.* Uma transição é mostrada em forma de arco, ou seja, que tem direção e vai de um nó a outro.



2.1.5 Processos de Markov

De…nição 2.4 *O processo de Markov é um tipo de processo estocástico onde somente o valor presente de uma variável é relevante para prever o futuro. A vantagem do processo de Markov é que ele simpli…ca a análise dos processos estocásticos.*

Geralmente, os preços das acções são modelados usando o Processo de Markov.

Exemplo 2.1 *Suponhamos que o preço de uma acção hoje custe 1.000$00. Se o preço desta acção segue um processo de Markov, as nossas previsões sobre a ‡utuação futura dos preços desta acção não devem levar em conta a ‡utuação ocorrida semana passada, mês passado ou no ano passado. a única informação importante quetemos para avaliar esta acção é seu preço actual, ou seja, 1.000$00. Normalmente os preços futuros são expressos em termos de distribuições de probabilidades. A propriedade de Markov nos diz que a distribnuição de probabilidades dos em qualquer tempo no futuro depende única e exclusivamente do preço actual da acção.*

A propriedade de Markov para o preço das acções é consistente com a *Forma Fraca da E…ciência de Mercado.*

² Forma Fraca da E…ciência de Mercado: Diz que o preço actual de uma acção já re‡ete plenamente todas as informações que estão contidas na sequencia histórica do preço. Daí segue que não existe nenhum benefício em se prever as movimentações futuras no preço das acções baseadas em séries históricas dos preços, ou seja, a propriedade de Markov se vere…ca.

Existem também outras duas formas de e…ciência de mercado, a *Forma Semi-Forte* e a *Forma Forte.*

² A Forma Semi-Forte diz que o preço actual da acção não somente re‡ete todas as informações contidas nos dados históricos, mas também re‡ete todos os conhecimentos públicos dispsponíveis.

² A Forma Forte vai além, dizendo que o preço actual re‡ete também todas as informações, sejam elas insider ou públicas.

Cadeias de Markov a Tempo Discreto

Exemplo 2.2 *Consideremos um jogo no qual em cada aposta um jogador perde 1.000$00 com probabilidade 0,6 ou o ganha com probabilidade 0,4.*

Suponhamos também que o mesmo jogador decide parar de jogar se a sua fortuna atingir N escudos e se ela atingir 0 escudos o jogador não pode continuar o jogo.

Seja X_n a quantidade de dinheiro que tem o jogador depois de n apostas. Observe que são esta quantidade e o resultado do próximo sorteio que vão determinar a sua fortuna depois da aposta seguinte. Qualquer que tenha sido a evolução da sua fortuna no passado (ou seja, os valores de $X_{n-1}, X_{n-2}, ..., X_0$), para prever o próximo estado X_{n+1}, é suficiente conhecer a sua fortuna no presente X_n. De facto, se $X_n = i$, com $0 < i < N$, então independentemente dos valores $i_0, ..., i_{n-1}$, teremos que:

$$P\left(X_{n+1} = i + 1.000\$00 \mid X_n = i,\ X_{n-1} = i_{n-1}, ...,\ X_0 = i_0\right) = 0,4$$

Pois isso significa que se o jogador ganhar a aposta $n + 1$ a sua fortuna vai ser acrescentada em mil escudos e portanto é suficiente conhecer o valor da sua fortuna no presente.

Em cada aposta a sua fortuna somente poderá aumentar ou diminuir em mil escudos com uma chance que não depende do número de aposta que ele já fez.

Imaginemos por exemplo, o valor $N = 5$ mil escudos. Então os valores que pode tomar a sua fortuna são

$$\{0 \text{ escudos, } 1 \text{ mil, } 2 \text{ mil, } 3 \text{ mil, } 4 \text{ mil, } 5 \text{ mil escudos}\}$$

Suponhamos, por exemplo, que depois de certa quantidade de apostas, o jogador tem 2 mil escudos. Podemos considerar as probabilidades de sua fortuna na próxima aposta, tomando algum destes valores possíveis. Como já foi observado, depois de uma aposta, o jogador terá ou 1 mil escudos ou 3 mil escudos, dependendo da sua sorte. As probabilidades mencionadas podem ser arranjadas em um vector linha da forma $(0;\ 0,6;\ 0;\ 0,4;\ 0)$. Visto que a soma destas probabilidades todas é um, pois estamos a considerar todos os valores possíveis da sua fortuna, ou seja, este vector corresponde a uma distribuição de probabilidade. Fazendo isto para cada valor possível da sua fortuna podemos arranjar os vectores de probabilidades como linhas de uma matriz que ficaria da forma seguinte:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0,6 & 0 & 0,4 & 0 & 0 & 0 \\ 0 & 0,6 & 0 & 0,4 & 0 & 0 \\ 0 & 0 & 0,6 & 0 & 0,4 & 0 \\ 0 & 0 & 0 & 0,6 & 0 & 0,4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

O exemplo acima corresponde a um tipo de processo estocástico muito importante, as chamadas cadeias de Markov.

Definição 2.5 (Cadeia de Markov) *Uma cadeia de Markov é um processo estocástico* $\{X_n\}_{n \in T}$, *com o tempo discreto,* $T = \{0, 1, 2, ...\}$, *o espaço de estados* E *finito ou enumerável e que possui a propriedade de Markov,*

$$P(X_{n+1} = j | X_n = i_n, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = P(X_{n+1} = j | X_n = i_n)$$

para todos os estados $i_0, ..., i_n, j$ e todo instante n. Se $X_n = i$ dizemos que o processo no instante n está no estado i.

A equação acima diz que o estado futuro do processo, $X_{n+1} = j$, não depende do passado, $X_{n-1} = i_{n-1}, ..., X_0 = i_0$, mas sim, só depende do presente, $X_n = i_n$.

A probabilidade condicional é chamada probabilidade de transição.

Para as cadeias de Markov homogêneas, isto é, aquelas cadeias nas quais não depende do tempo n,

$$P(X_{n+1} = j | X_n = i,) = ... = P(X_1 = j | X_0 = i) = P_{i,j} \qquad i, j \in E,$$

logo $P_{i,j}$ é a probabilidade de passar, em qualquer instante, do estado i ao estado j.

Podemos escrever as probabilidades de transição $P_{i,j}$ numa matriz P, como foi feito no exemplo anterior, que é chamada de matriz de transição.

Se E é finito, por exemplo $E = \{0, 1, ..., N\}$, então:

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,N} \\ P_{1,0} & P_{1,1} & \cdots & P_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,1} & \cdots & P_{N,N} \end{bmatrix}$$

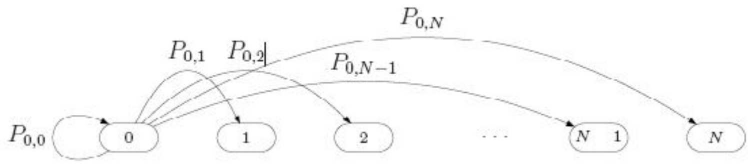
No caso do exemplo do jogador, as probabilidades de transição não nulas são:

$$\begin{aligned} P_{i,i+1} &= 0,4;\ P_{i,i-1} = 0,6;\ \text{se } 0 < i < N \\ P_{0,0} &= 1 = P_{N,N}. \end{aligned}$$

Se E é infinito, por exemplo $E = \{0,1,2,\ldots\}$ então a matriz P será infinita:

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots \\ P_{2,0} & P_{2,1} & P_{2,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Também podemos escrever as transições como um diagrama:



Cadeias de Markov a Tempo Continuo

Os modelos discretos são úteis para responder as questões pontuais, no que diz respeito ao tempo, entretanto, quando se deseja saber, por exemplo, qual o tempo esperado exacto do reparo de uma máquina ou o tempo exacto até que duas máquinas em funcionamento sejam desligadas para reparo, o modelo discreto não é satisfatório. Assim, faz-se necessário a inclusão de um processo estocástico de tempo contínuo.

Definição 2.6 *Consideremos um processo estocástico a tempo contínuo $X = \{X_t\}_{t \geq 0}$ com espaço de estado E finito ou enumerável. Diremos que X é uma cadeia de Markov a tempo contínuo se e somente se para todo $t,\ s \geq 0$ tem-se:*

$$P(X_{t+s} = j | X_u,\ u \leq s) = P(X_{t+s} = j | X_s).$$

Se além disto, a probabilidade de transição entre dois estados depende somente de intervalo de tempo durante o qual ocorre a transição e não dos instantes de tempo nos que a cadeia ocupa esses estados, ou seja, quando:

$$P\left(X_{t+s}=j \mid X_s=i\right)=P_{i,j}(t),$$

a cadeia será chamada de Homogênea no tempo.

2.1.6 O Processo de Nascimento e Morte

A maioria dos modelos elementares de filas de espera assume que as entradas - chegadas de novos clientes e as saídas - saídas de clientes do sistema de fila de espera ocorre de acordo com o *processo de nascimento e morte*, em que:

— Nascimento: chegada de clientes ao sistema.

— Morte: saída de cliente (com serviço concluído) do sistema.

Definição 2.7 *Um processo puro de nascimento é aquele onde só ocorrem nascimentos e nenhuma morte; um processo puro de morte é aquele onde só ocorrem mortes e nunhum nascimento.*

Exemplo 2.3 *Uma faculdade anunciou uma vaga para o lugar de director académico, tendo estipulado uma data limite até à qual seriam recebidas candidaturas ao lugar. Se não for levado a cabo nenhum processamento das canditaturas até à data limite e se não houver adesistência de qualquer dos candidatos inscritos, então o processo de recpção é um processo puro de nascimento até à data limite. Se nenhuma inscrição for aceite depois da data limite, então o processo de redução do conjunto de candidaturas em análise por avaliação e eliminação é um processo puro de morte. Se as inscrições são processadas durante o período em que ainda são aceites candidaturas, o processo é um processo de nascimento-morte. Em qualquer dos casos, a população é o conjunto de candidaturas em análise.*

Processo Markoviano Generalizado de Nascimento-Morte

Definição 2.8 *Um processo de crescimento populacional é um processo de Markov se as probabilidades de transição para mudança de um estado para outro dependerem apenas do estado corrente e não da história passada recente do processo até chegar ao estado corrente. Mais formalmente, um processo markoviano generalizado de nascimente-morte satisfaz os seguintes critérios:*

— As distribuições de probabilidade que governam os números de nascimentos e mortes num determinado intervalo de tempo dependem da amplitude do intervalo,mas não do seu instante inicial.

— A probabilidade de haver exactamente um nascimento num intervalo de tempo de amplitude Δt, dada uma população de n membros no começo do intervalo é $\lambda_n \Delta t + o(\Delta t)$, onde λ_n é uma constante, possivelmente diferente para diferentes valores de n.

— A probabilidade de haver exactamente uma morte num intervalo de tempo de amplitude Δt, dada uma população de n membros no começo do intervalo é $\mu_n \Delta t + o(\Delta t)$, onde μ_n é uma constante, possivelmente diferente para diferentes valores de n.

— A probabilidade de haver mais de um nascimente e a probabilidade de haver mais de uma mortenum intervalo de tempo de amplitude Δt são ambas iguais a $o(\Delta t)$.

Estes critérios implicam, no limite quando Δt tende para zero, *as equações de Kolmogorov* para as probabilidades de estado:

$$\frac{dp_n(t)}{dt} = -(\lambda_n + \mu_n)\, p_n(t) + \mu_{n+1} p_{n+1}(t) + \lambda_{n-1} p_{n-1}(t) \qquad (n = 1, 2, \ldots)$$

$$\frac{dp_0(t)}{dt} = -\lambda_0 p_0(t) + \mu_1 p_1(t)$$

Exercício 2.2 *Deduza as equações de Kolmogorov.*

Resolução:

O tamanho da população no instante de tempo $t + \Delta t$, $N(t + \Delta t)$ é dependente do tamanho no instante de tempo t, $N(t)$, bem como de quaisquer mudanças (nascimentos e/ou mortes) ocorridas no intervalo de tempo $]t,\ t + \Delta t]$. Assim, para $n > 1$,

$P\{N(t + \Delta t) = n\} = P\{N(t) = n$ e ocorrem 0 nascimentos e 0 mortes em $]t,\ t + \Delta t]\} + P\{N(t) = n$ e existem 1 nascimento e 1 morte em $]t,\ t + \Delta t]\} + P\{N(t) = n - 1$ e ocorrem 1 nascimento e 0 mortes em $]t,\ t + \Delta t]\} + P\{N(t) = n + 1$ e ocorrem 0 nascimentos e 1 morte em $]t,\ t + \Delta t]\} + P\{$uma combinação adequada de acontecimentos envolvendo mais de 1 nascimento ou mais de 1 morte em $]t,\ t + \Delta t]\}$

ou

$$p_n(t + \Delta t) = a + b + c + d + e$$

Utilizando a noção de probabilidade condicionada, temos $a = P\{N(t) = n\} \times P\{0$ nascimento e 0 mortes em $\Delta t \mid N(t) = n\}$

Das hipóteses de base, a probabilidade de zero nascimentos num intervalo de tempo de amplitude Δt é, a menos de $o(\Delta t)$, 1 menos a probabilidade de ocorrer exactamente um nascimento; dado o estado n no começo do intervalo, esta última probabilidade é igual a $\lambda_n \Delta t + o(\Delta t)$. Assim, a probabilidadede zero nascimentos é

$$1 - \lambda_n \Delta t + o(\Delta t)$$

e, sob as mesmas condições, a probabildade de zero mortes é

$$1 - \mu_n \Delta t + o(\Delta t)$$

Além disso, como os nascimentos ocorrem independentemente de mortes,

$$\begin{aligned} a &= p_n(t) \times [1 - \lambda_n \Delta t + o(\Delta t)][1 - \mu_n \Delta t + o(\Delta t)] = \\ &= p_n(t) \times [1 - (\lambda_n + \mu_n) \Delta t + o(\Delta t)] \end{aligned}$$

Do mesmo modo, obtemos:

$$b = o(\Delta t)$$

$$c = p_{n-1}(t)(\lambda_{n-1} \Delta t) + o(\Delta t)$$

$$d = p_{n+1}(t)(\lambda_{n+1} \Delta t) + o(\Delta t)$$

$$e = o(\Delta t)$$

Portanto,

$$p_n(t+\Delta t) = p_n(t) + \left[-(\lambda_{n-1} + \mu_n)p_n(t) + \mu_{n-1}p_{n-1}(t) + \mu_{n+1}p_{n+1}(t)\right]\Delta t + o(\Delta t)$$

Passando $p_n(t)$ para o primeiro membro da equação, dividindo por Δt e fazendo $\Delta t \to 0$, obtemos as equações de Kolmogorov para $n = 1, 2, \ldots$.

O caso $n = 0$ precisa de ser analisado separadamente, já que não são possíveis mortes no estado 0. Levando a cabo a análise apresentada acima, rapidamente obtemos a equação de Kolmogorov para o caso $n = 0$.

Processo Markoviano Linear de Nascimento

Definição 2.9 *Um processo markoviano linear de nascimento é um processo markoviano puro de nascimento no qual a probabilidade de haver um nascimentonum curto intervalo de tempo é proporcional querao número actualde membros na população, quer à amplitude do intervalo.Isto é, para todo o* n, $\mu_n = 0$ e $\lambda_n = n\lambda$. *A constante de proporcionalidade* λ *é a taxa de nascimento ou a taxa de chegada. Para uma população inicial de um membro, é:*

$$p_n(t) = \begin{cases} \left(1 - e^{-\lambda t}\right)^{n-1} e^{-\lambda t} & (n = 1, 2, \ldots) \\ 0 & (n = 0) \end{cases}$$

O tamanho esperado da população no instante de tempo t é $E[N(t)] = e^{\lambda t}$. Se a população inicial é de $N(0)$ membros, então o tamanho esperado no instante de tempo t é $E[N(t)] = N(0)\, e^{\lambda t}$.

Exercício 2.3 *Um processo markoviano linear de nascimento iniciado com um membro tem uma taxa média horária de nascimento de* $\lambda = 2$. *Determine a possibilidade de se ter uma população maior que* 3 *após* 1 *hora e o tamanho esperado da população nessa altura.*

Resolução:

com $\lambda = 2$ novos nascimentos por membro por hora e com $t = 1h$, obtem-se:

$$p_0(1) = 0 \qquad p_1(1) = \left(1 - e^{-2}\right)^0 e^{-2} = 0,135$$

$$p_2(1) = \left(1 - e^{-2}\right)^1 e^{-2} = 0,117 \qquad p_3(1) = \left(1 - e^{-2}\right)^2 e^{-2} = 0,101$$

A probabilidade de haver mais de três membros na população após 1 hora é então

$$1 - (0 + 0,135 + 0,117 + 0,101) = 0,647$$

E o tamanho esperado da população nessa altura é

$$E[N(1)] = 1e^{2(1)} = 7,389 \text{ membros.}$$

Processo Markoviano linear de Morte

Definição 2.10 *Um processo markoviano linear de morte é um processo markoviano puro de morte no qual a probabilidade de haver uma morte num curto intervalo de tempo é proporcional quer ao tamanho actual da população quer à amplitude do intervalo. Isto é, para todo o* n, $\lambda_n = 0$ *e* $\mu_n = n\mu$. *A constante de proporcionalidade* μ *é a taxa de mortalidade.*

Assim, para uma população inicial de $N(0)$, é:

$$p_n(t) = \begin{cases} \binom{N(0)}{n} e^{-n\mu t}(1 - e^{-\mu t})^{N(0)-n} & [n \leq N(0)] \\ 0 & [n > N(0)] \end{cases}$$

E o tamanho esperado para a população no instante de tempo t é $E[N(t)] = N(0)e^{-\mu t}$.

Exercício 2.4 *Um processo markoviano linear de morte iniciado com* 10 *membros tem uma taxa média semanal de mortes de* $\mu = 0,6$. *determine a probabilidade de haver uma população de, pelo menos,* 8 *membros após* 3 *dias e o tamanho esperado da população nessa altura.*

Resolução:

Com $N(0) = 10$, $t = \left(\frac{3}{7}\right)$ $semana$ e $\mu = 0,6$ mortes por membros por semana, obtém-se:

$$p_8(3/7) = \binom{10}{8} e^{-8(0,6)(3/7)}\left(1 - e^{-(0,6)(3/7)}\right)^{10-8} = 45(0,1278)(1-0,7733)^2 = 0,296$$

$$p_9(3/7) = \binom{10}{9} e^{-9(0,6)(3/7)}\left(1 - e^{-(0,6)(3/7)}\right)^{10-9} = 10(0,0988)(1-0,7733)^1 = 0,224$$

$$p_{10}(3/7) = \binom{10}{10} e^{-10(0,6)(3/7)}\left(1 - e^{-(0,6)(3/7)}\right)^{10-10} = 1(0,0764)(1-0,7733)^0 = 0,076$$

Assim, a probabilidade de haver oito ou mais membros na população após três dias é:

$$p_8(3/7) + p_9(3/7) + p_{10}(3/7) = 0,596$$

E o tamanho esperado da população nessa altura é

$$E[N(3/7)] = 10e^{-(0,6)(3/7)} = 7,73 \text{ membros}.$$

Processo Markoviano Linear de Nascimento-Morte

Definição 2.11 *Um processo markoviano linear de nascimento-morte é um processo markoviano de nascimento-morte no qual, para todo n, $\lambda_n = n\lambda$ e $\mu_n = n\mu$.*

Assim, para uma população inicial de um membro, tem-se:

$$p_n(t) = \begin{cases} [1 - r(t)][1 - s(t)][s(t)]^{n-1} & (n = 1, 2, ...) \\ r(t) & (n = 0) \end{cases}$$

onde,

$$r(t) \equiv \frac{\mu\left[e^{(\lambda-\mu)t} - 1\right]}{\lambda e^{(\lambda-\mu)t} - \mu} \quad \text{e} \quad s(t) \equiv \frac{\lambda\left[e^{(\lambda-\mu)t} - 1\right]}{\lambda e^{(\lambda-\mu)t} - \mu}$$

E o tamanho esperado da população no instante de tempo t é

$$E[N(t)] = e^{(\lambda-\mu)t}.$$

Se a população inicial é de $N(0)$ membros, então o tamanho esperado no instante de tempo é

$$E[N(t)] = N(0)e^{(\lambda-\mu)t}.$$

Exercício 2.5 *Um Biólogo observa o crescimento de uma dada bactéria em cultura e conclui que quer a probabilidade de nascimento quer a probabilidade de morte de uma bactéria são proporcionais ao número de bactérias em cultura e ao tempo decorrido. Em média, cada bactéria origina uma nova em cada 7 horas e morre após 30 horas. Quantas bactérias se podem esperar numa cultura após uma semana, se a população é iniciada com uma bactéria?*

Resolução:

Considerando um dia como unidade de tempo, temos $N(0) = 1$,
$\lambda = \left(\frac{1}{7}\right)(24) = 3,43$ nascimentos por membro por dia;
e $\mu = \left(\frac{1}{30}\right)(24) = 0,8$ mortes por membro por dia.

Portanto, O tamanho esperado da população após uma semana (7 dias) é

$$E[N(7)] = 1e^{(3{,}43-0{,}8)7} = 97.953.164 \text{ bactérias.}$$

Nota 2.3 *Um processo markoviano linear de nascimento-morte com parâmetro λ e μ e uma população inicial $N(0)$ é equivalente à soma de $N(0)$ processos simultâneos mas independentes, cada um com parâmetros λ e μ e uma população inicial igual a 1.*

Hipóteses do Processo de Nascimento e Morte

Sendo $N(t)$ – o número de clientes no sistema e dado $N(t) = n$, então a distribuição de probabilidade do tempo até à próxima chegada é uma exponencial de parâmetro λ_n e a distribuição de probabilidade do tempo até à próxima saída é uma exponencial de parâmetro μ_n. As variáveis aleatórias tempo até à próxima chegada e tempo até à próxima saída são mutuamente independentes.

As transições de estado, isto é, as alterações do número de clientes no sistema são:

- $n \to n + 1$ (um único nascimento);
- $n \to n - 1$ (uma única morte).

Exemplo 2.4 *Formule o processo seguinte como uma cadeia de markov e construa o diagrama de transição de estados:*

"O programa de formação de supervisores de produção de uma determinada companhia consiste em duas fases. A fase 1, a qual envolve três semanas de aulas, é seguida da fase 2, que envolve três semanas de aprendizagem prática sob a direcção de supervisores experimentados. Da experiência anterior, a companhia espera que 60% dos candidatos que começam as aulas venham a passar à fase seguinte, verificando-se a desistência dos restantes 40%. Dos que passam à fase prática, 70% serão graduados como supervisores, 10% terão de repitir esta fase e 20% abandonarão o programa."

Quantos supervisores pode a companhia esperar formar no seu programa de formação actual, sabendo que há 45 pessoas na primeira fase e 21 na fase prática?

Resolução:

Cadeia de Markov

Consideramos um período de tempo com 3 semanas e definimos os estados de 1 a 4 correspondentes às seguintes condições: desistências, fase de aulas, fase de apredizagem prática e graduação como supervisor, repectivamente. Se assumirmos que os candidatos eliminados não voltam ao programa de formação e que os supervisores permanecem como supervisores, então as probabilidades de transição são dadas pela seguinte matriz estocástica

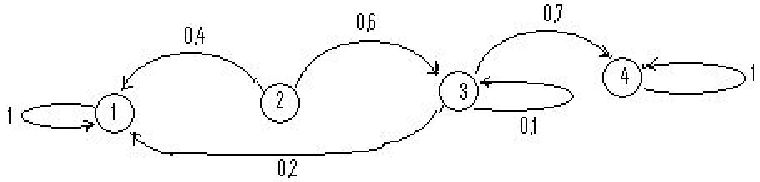
$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0,4 & 0 & 0,6 & 0 \\ 0,2 & 0 & 0,1 & 0,7 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Há (45 + 21) pessoas =66 pessoas no programa de formação actual, pelo que o vector de probabilidade inicial é

$$X^{(0)} = \begin{bmatrix} 0, & \frac{45}{66}, & \frac{21}{66}, & 0 \end{bmatrix}$$

Diagrama de transição de estados

Temos o diagrama de transição de estados representado na figura abaixo, onde o número em cada arco é a correspondente probabilidade de transição:



CAPÍTULO 3

Desempenho dos Modelos

Existem os seguintes modelos: Markovianos, Não-Markovianos e em Série.

Os modelos em Série são: Modelos de Rede Aberta e Modelos de Rede Fechada. Entretanto, vamos começar a falar da famosa Lei de Little.

3.1 A Lei de Little

Um dos mais poderosos relacionamentos em teoria das …las foi desenvolvido por John Little no início dos anos 60. Little relacionou o tamanho médio do sistema em estado de equilírio com o tempo médio de espera dos clientes, também em estado de equilíbrio, da seguinte forma. Uma vez que T_q é o tempo que um cliente gasta esperando na …la antes de entrar em serviço e T o tempo total que um cliente gasta no sistema ($T = T_q + S$, onde S é o tempo de serviço, e T, T_q e S são variáveis aleatórias).

A Fórmula de Little é uma analogia com a fórmula de Física Fundamental, isto é, distância = velocidade £ tempo. Ou seja, os princípios físicos são aplicados no sistema de …las de espera.

Portanto,

O número médio de clientes no sistema de …la = taxa de chegada £ tempo médio no sistema.

$$L = \lambda £ W$$

Apesar de ser uma lei intuitivamente razoável, trata-se de um resultado excepcional, se considerarmos que é válido para a grande maioria das situações, não sendo necessário nenhum conhecimento adicional sobre os processos de chegada e saída. A única exigência é que o número de clientes que chegam ao sistema seja igual ao número de clientes que têm seus serviços completados.

Podemos obter a partir da Lei de Little os seguintes relacionamentos:

Número médio de clientes na …la de espera = taxa de chegada £ tempo médio na …la

$$L_q = \lambda \text{ £ } W_q$$

Número médio de clientes em atendimento = taxa de chegada £ tempo médio de atendimento

$$L_s = \lambda \text{ £ } W_s$$

Outra caracteristica interessante da Lei de Little é o facto de ela poder ser aplicada nos mais diferentes níveis de um sistema, desde o nível recurso (centro de serviço) até o nível de todo o sistema.

Exemplo 3.1 *Suponhamos que um disco tenha sido monitorado por um tempo e que se tenha veri…cado que a taxa de chegada seja de* 150 *requisições por segundo e que o tempo médio de serviço seja de* 10 *milisegundos. É possível determinar o número médio de requisições no disco como:*

$$L = 150 \text{ £ } 0,01 = 1,5 \text{ requisições}$$

Exercício 3.1 *Deduza a lei de Little intuitivamente.*

Resolução:

Durante o tempo de permanência do "cliente médio" no sistema, W, novos clientes chegam a uma taxa média λ; Assim, no …m de W unidades de tempo, λW novos clintes são esperados no sistema. Isto é, quando o cliente inicial deixa o sistema, ele pode esperar ver λW outros clientes que permanecem no sistema. Como as estatísticas da …la de espera são independentes do tempo no estado estacionário, veri…ca-se sempre $L = \lambda W$.

3.2 Modelos Markovianos

Quando os tempos de chegada e atendimento têm probabilidade com distribuição exponencial, o sistema de …las é um sistema de …las Markoviano. A distribuição exponencial satisfaz a hipótese de Markovian na qual ela não tem memória. Ou seja, ao esperar a ocorrência de uma chegada com o tempo distribuído exponencialmente, o valor esperado do tempo de chegada não depende de quanto tempo já transcorreu, esperando essa chegada. Esta situação mostra-se peculiar, mas existem situações em que esta hipótese é valida. E a suposição da distribuíção exponencial é necessária para obter soluções aproximadas para variáveis estatisticas. Com a hipótese de distribuicções exponenciais, os processos de chegada e atendimento são *processos de Poisson.*

Todos os modelos markovianos podem ser analisados como um processo de nascimento e morte.

São modelos Markovianos:

- Modelo $M/M/1/GD/$ 1 / 1
- Modelo $M/M/1/GD/c/$ 1
- Modelo $M/M/S/GD/$ 1 / 1
- Modelo $M/M/R/GD/K/K$ de reparo de máquinas
- Modelos $M/G/$ 1 $/GD/$ 1 / 1 e $GI/G/$ 1 $/GD/$ 1 / 1 /

3.3 Modelos não Markovianos

Resultados analíticos disponíveis somente existem em poucas situações em que não é necessária a hipótese de Markov nos processos de chegada e/ou atendimento. Eles são importantes porque, em situações práticas, as distribuições de tempo de chegada e atendimento não poderão ser razoavelmente aproximadas por distribuições exponenciais. Para sistemas de …las com modelos $M/G/1$, o resultado das fórmulas são exactos. Para os demais modelos de sistemas de …las, os resultados das fórmulas são aproximados. Não existem fórmulas disponíveis para …las de espera …mitas ou populações …nitas num sistema de …las Não-Markovianos.

Os sistemas Não-Markovianos precisam de especi…cação dos coe…cientes de variação (COV) do processo de chegada e atendimento. O COV de chegadas é o resultado do desvio padrão de tempo entre chegadas dividido pelo o valor esperado do tempo entre chegadas. O COV de uma distribuição exponencial é 1. Distribuições com menor variabilidade que a distribuição exponencial têm o COV < 1, enquanto que distribuições com maior variabilidade que a distribuição exponencial têm o COV > 1. O COV do processo de atendimento é o desvio padrão do tempo de atendimento dividido pelo valor esperado do tempo de atendimento.

São Modelos Não-Markovianos:

² Modelo $M/G/1/GD/$ 1 / 1

² Modelo $M/G/S/GD/S/$ 1

² Modelo $G/G/M$

3.4 Modelos em Série

Em muitas situações, uma unidade de chegada não somente passa por uma …la de espera, mas por uma série de …las de espera. Um exemplo é matricular-se em uma universidade. Estudante visitam um número de departamentos requeridos para a aprovação do programa acadêmico. Cada departamento tem um ou mais atendentes e uma …la de espera para os estudantes. Outro exemplo é em uma fábrica onde os equipamentos são agrupados de acordo com a sua função em diferentes estações. Um grupo de equipamentos similares pode ser modelado como uma estação de um sistema de …las. A unidade de chegada passa por uma ou mais estações antes de sair da fábrica. A análise deste tipo de redes pode ser sumamente difícil a menos que eles tenham uma estrutura especial.

As redes de …las de espera são sistemas dinâmicos constituídos por entidades processadoras pelas quais ‡uem clientes que aguardam a sua vez em …la. Depois de servidos numa dada entidade, os clientes deslocam-se para uma nova …la de espera ou abandonam o sistema.

3.4.1 Rede de Fila de Espera Aberta

A rede de …la de espera aberta é uma generalização do modelo de …las de espera em série.

No entanto as razões de chegada podem ser encontradas resolvendo um sistema de equações da seguinte maneira: Cada estação tem uma equação. Por exemplo, duas estações teriam duas equações com duas incógnitas, três estações teriam três equações com três incógnitas, etc.

O sistema de equações geral é o seguinte:

$$\lambda_j = r_j + \sum_{j=1}^{j=k} \lambda_i P_{i,j}.$$

Onde, j representa uma estação qualquer;
r_j = razão de chegada externa à estação j;
λ_i = razão de chegada à estação i;
$P_{i,j}$ = probabilidade de transição da estação i à estação j.

3.4.2 Rede de Fila de Espera Fechada

Em fábricas, indústrias e sistemas de computação, onde existem um constante número de diversas actividades acontecento. Estes podem ser modelados como uma rede de …la de espera fechada. Quando um trabalho (cliente) termina de ser atendido (sai do sistema), ele é substituído por outro que entra no sistema, assim, sempre mantendo constante o número de clientes no sistema. Lembrando que nas redes em …la de espera aberta, o número de actividades em cada atendimento eram variáveis aleatórias independentes. Aqui, como o número de trabalhos (clientes) no sistema é sempre constante, a distribuição de actividades (clientes) nos diferentes servidores (atendentes) não pode ser independente.

3.5 Comportamento Transitório de um Sistema de Filas

Por tudo dito até agora, a razão de chegada, a razão de atendimento e o número de atendentes são constantes através do tempo. Isto permite falar da existência de um estado estável. Mas em muitas situações as razões de chegada e atendimento, assim como o número de atendentes, podem variar. Quando os parâmetros que de…nem o sistema de uma …la de espera variam através do tempo, chamamos esse sistema de *não estacionário* ($nonstationary$).

3.5.1 Probabilidades Transitórias

Consideremos um exemplo de um restaurante de comida rápida (fast-food) que abre às 10 da manhã e fecha às 18 horas. A distribuição de probabilidade do número de clientes presentes a todo o momento entre as 10 e as 18 são chamadas probabilidades transitórias. Por exemplo, para determinar a probabilidade de que pelo menos dez clientes estejam presentes, esta probabilidade certamente será maior às 12:30 PM que às 15 horas.

3.5.2 Processo Não Homogêneo de Poisson

Neste processo acontecem três factos:

1. Num instante qualquer t, a probabilidade de uma chegada num sistema de ...las de espera é

$$\lambda(t) \text{ £ } ¢t.$$

2. No instante $t,$ a probabilidade de chegada de mais de um cliente é $o($¢$t)$.

3. Chegadas durante intervalos diferentes são independentes (a razão de chegada pode variar em cada intervalo).

CAPÍTULO 4

Considerações Finais

É com uma grande satisfação que chegamos ao …m deste trabalho, apesar dos obstáculos encontrados (complexidade do tema e alguma carência do material didáctico), temos que salientar que este trabalho fez com que alargassemos os nossos conhecimento a nível do tema em questão.

Esperamos ter alcançado os objectivos traçados para o estudo do tema. Ainda, esperamos que este trabalho venha a servir como um instrumento de pesquisa para todos aqueles que queiram iniciar os seus estudos sobre a Teoria das Filas de Espera.

Uma sugestão que gostaríamos de deixar aqui é que o Departamento de Matemática pensasse numa forma de introduzir este tema no programa curricular do curso, como uma ponte de interligação das disciplinas de Probabilidade, Estatistica, Álgebra e Análise.

CAPÍTULO 5

Bibliogra...a

[1] Bronson, Richard e Naadimuthu, Govindasami "Investigação Operacional" Segunda Edição Mc Graw Hill.

[2] Karlin, samuel e Taylor, Howard " A First Course in Stochastic Processes" Second Edition Academic Press, 1975.

[3] Lefebvre, Mario "Applied Stochastic Processes" Springer.

[4] Reis, Elizabeth "Estatística Descritiva" Edições Sílabo, Lda.

[5] Neves, Maria Augusta Ferreira e Fernandes, José António "Métodos Quantitativos" Ensino Secundário. Porto Editora.

CAPÍTULO 6

Anexo

6.1 Notação

Certos sistemas de ﬁlas ocorrem tão frequentemente na prática, que as notações padrões têm sido desenvolvidas para eles. Uma notação populararmente usada é a notação de Kendall (1951). É traduzida por uma terminologia que envolve 6 elementos, compondo o seguinte padrão: $1/2/3/4/5/6$.

A primeira característica especiﬁca a natureza do processo de chegada, onde as seguintes abreviações são usadas:

M = Chegadas com tempos de chegadas independentes, variáveis randômicas identicamente distribuídas (iid) descritas por distribuições exponenciais. Ou chegadas por unidade de tempo independentes, variáveis aleatórias identicamente distribuídas (iid) descritas por distribuições de Poisson.

D = Chegadas com tempo de chegada são iid e determinísticas (variância = 0).

E_k = Chegadas com tempo de chegada são iid e descritas pela distribuição de Erlang de parâmetro

$$R = \lambda k$$

e com forma k.

GI = Chegadas com tempo de chegada são iid e controladas por uma distribuição geral.

A segunda característica especiﬁca a natureza dos tempos de atendimento:

M = Tempos de atendimento independentes, variáveis randômicas identicamente distribuídas (iid) e distribuídas exponencialmente.

D = Tempos de atendimento iid e determinísticas.

E_k= Tempos de atendimento são iid descritas pela distribuição de Erlang de parâmetro

$$R = \lambda k$$

com forma k.

G= Tempos de atendimento são iid e seguindo uma distribuição geral.

A terceira característica é o número de atendentes em paralelo.

A quarta característica descreve a disciplina da …la de espera:

$FCFS$ = primeiro a chegar, primeiro a ser atendido (Atendimento por ordem de chegada).

$LCFS$ = último a chegar, primeiro em ser atendido (Atendimento por ordem inversa da de chegada).

$SIRO$ = Atendimento em ordem randômico (Atendimento por ordem aleatória).

PRI = Atendimento por ordem com prioridades.

GD = Disciplina geral de uma …la de espera.

A quinta característica denota o máximo número permitido de clientes no sistema, incluindo clientes que estão na …la de espera e clientes que estão sendo atendidos pelo servidor.

A sexta característica signi…ca o tamanho da população. A população geralmente é considerada in…nita.

Exemplo 6.1 *Um sistema $M/E_2/8/FCFS/10/$ 1 pode representar uma clínica de saúde com oito doutores, tempos de chegada exponenciais, tempo de atendimento Erlang com duas faces, a disciplina da …la de espera é $FCFS$ (o primeiro em chegar é atendido) e uma capacidade total de dez pacientes.*

Nota 6.1 *Em muitos modelos importantes, as três últimas variáveis são omitidas quando seus valores respectivos são GD/ 1 / 1 .*

Exemplo 6.2 *Um sistema $M/M/1$ tem chegadas com distribuição de Poisson, atendimento com distribuição exponencial, um atendente, disciplina geral, in…nito número de clientes no sistema e in…nita população. Aqui as três últimas variáveis são omitidas mas com signi…cado implícito.*